```
1  TACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGGA
   ATGATTTCCC TTGTTTTCGA CCTCGAGGTG GCGCCACCGC CGGCGAGATC TTGATCACCT
                                        5' UTR
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

   TCCCCCGGGC TGCAGGAATT CGGCACGAGG AACTTTCTGC CTCGTTTTTT TGCTCCTACT
   AGGGGGCCCG ACGTCCTTAA GCCGTGCTCC TTGAAAGACG GAGCAAAAAA ACGAGGATGA
      5' UTR                                SEQ ID NO: 3
   ~~~~~~~~~~~~~~~~~~~~~~~~~   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~


                          M  S  Q  E  I  V  Q  S  G  Q  T  Y
   GTTTTTCTCT TCCAGTTTCT ACCATGTCGC AAGAAATTGT TCAATCAGGA CAAACCTACA
   CAAAAAGAGA AGGTCAAAGA TGGTACAGCG TTCTTTAACA AGTTAGTCCT GTTTGGATGT
      SEQ ID NO: 3
   ~~~~~~~~~~~~~~~~~~
   I  I  T  N  A  K  S  G  T  V  V  D  L  S  G  E  D  N  K  S
   TCATCACTAA CGCCAAATCC GGCACAGTTG TTGACCTTTC GGGCGAAGAC AACAAATCTA
   AGTAGTGATT GCGGTTTAGG CCGTGTCAAC AACTGGAAAG CCCGCTTCTG TTGTTTAGAT

   I  I  G  F  P  K  H  G  G  T  N  Q  R  W  T  L  N  W  T  G
   TTATTGGATT TCCCAAGCAT GGAGGAACAA ATCAGAGGTG GACCCTCAAC TGGACAGGGA
   AATAACCTAA AGGGTTCGTA CCTCCTTGTT TAGTCTCCAC CTGGGAGTTG ACCTGTCCCT
                                        SEQ ID NO: 5
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~


   K  S  W  T  F  R  S  V  S  S  E  M  Y  L  G  L  N  G  S  P
   AGAGTTGGAC TTTCCGCTCC GTTTCTTCTG AAATGTATCT TGGCCTGAAT GGCTCGCCGT
   TCTCAACCTG AAAGGCGAGG CAAAGAAGAC TTTACATAGA ACCGGACTTA CCGAGCGGCA
                          SEQ ID NO: 4 (partial)
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      SEQ ID NO: 5
   ~~~~~~~~~~~~~~~~~
                          SEQ ID NO: 6 (partial)
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   S  D  G  T  K  L  V  A  V  T  P  V  E  W  R  I  W  H
   CTGATGGAAC AAAACTGGTA GCCGTGACCA CCCCTGTTGA GTGGCGCATC TGGCACGA    418
   GACTACCTTG TTTTGACCAT CGGCACTGGT GGGGACAACT CACCGCGTAG ACCGTGCT
```

FIG. 1

```
                      5' UTR
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                              START
                                              ~~~
                                              M  S   Q  E  I
  1   GCCTCGTTTT TTTGCTCCTA CTGTTTTTCT CTTCCAGTTT CTACCATGTC GCAAGAAATT
      CGGAGCAAAA AAACGAGGAT GACAAAAAGA GAAGGTCAAA GATGGTACAG CGTTCTTTAA

      V  Q  S   G  Q  T  Y   I  I  T   N  A  K   S  G  T  V   V  D  L
 61   GTTCAATCAG GACAAACCTA CATCATCACT AACGCCAAAT CCGGCACAGT TGTTGACCTT
      CAAGTTAGTC CTGTTTGGAT GTAGTAGTGA TTGCGGTTTA GGCCGTGTCA ACAACTGGAA

      S  G  E   D  N  K  S   I  I  G   F  P  K   H  G  G  T   N  Q  R
121   TCGGGCGAAG ACAACAAATC TATTATTGGA TTTCCCAAGC ATGGAGGAAC AAATCAGAGG
      AGCCCGCTTC TGTTGTTTAG ATAATAACCT AAAGGGTTCG TACCTCCTTG TTTAGTCTCC

      W  T  L   N  W  T  G   K  S  W
181   TGGACCCTCA ACTGGACAGG GAAGAGTTGG A    211
      ACCTGGGAGT TGACCTGTCC CTTCTCAACC T
```

FIG. 2

```
         V   D   L   S   G   E   D   N   K   S   I   I   G   F   P   K   H   G   G
    1  TTGTTGACCT TTCGGGCGAA GACAACAAAT CTATTATTGG ATTTCCCAAG CATGGAGGAA
       AACAACTGGA AAGCCCGCTT CTGTTGTTTA GATAATAACC TAAAGGGTTC GTACCTCCTT


         T   N   Q   R   W   T   L   N   W   T   G   K   S   W   T   F   R   S   V   S
   61  CAAATCAGAG GTGGACCCTC AACTGGACAG GGAAGAGTTG GACTTTCCGC TCCGTTTCTT
       GTTTAGTCTC CACCTGGGAG TTGACCTGTC CCTTCTCAAC CTGAAAGGCG AGGCAAAGAA


         S   E   M   Y   L   G   L   N   G   S   P   S   D   G   T   K   L   V   A   V
  121  CTGAAATGTA TCTTGGCCTG AATGGCTCGC CGTCTGATGG AACAAAACTG GTAGCCGTGA
       GACTTTACAT AGAACCGGAC TTACCGAGCG GCAGACTACC TTGTTTTGAC CATCGGCACT


         T   T   P   V   E   W   H   I   W   H   D   E   V   D   P   S   T   Y   R   I
  181  CCACCCCTGT TGAGTGGCAC ATCTGGCACG ACGAAGTTGA CCCTTCAACT TATCGTATCT
       GGTGGGGACA ACTCACCGTG TAGACCGTGC TGCTTCAACT GGGAAGTTGA ATAGCATAGA

                                                          A/G polymorphism
                                                             ~

         F   V   P   F   T   T   F   N   M   D   L   Y   A   Q   G   S   A   P   G
  241  TTGTACCTTT CACCACATTC AACATGGACC TCTACGCCCA RGGTAGTGCC GCCCCTGGTA
       AACATGGAAA GTGGTGTAAG TTGTACCTGG AGATGCGGGT YCCATCACGG CGGGGACCAT

                                  T/C polymorphism
                                         ~

         T   P   I   T   T   W   Y   T   W   K   G   I   H   Q   T   W   R   F   E   L
  301  CGCCTATCAC AACTTGGTAT ACATGGAAGG GYATCCACCA AACGTGGAGG TTTGAACTAG
       GCGGATAGTG TTGAACCATA TGTACCTTCC CRTAGGTGGT TTGCACCTCC AAACTTGATC

         T/G polymorphism
             ~
         STOP
         ~~~
                                         3' UTR
         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         A  *
  361  CTTAGGKTCA GGTTTCGGAT GTAATTTGTG TGTGTAAATC TTCTTGGACC ATGTTGTGCT
       GAATCCMAGT CCAAAGCCTA CATTAAACAC ACACATTTAG AAGAACCTGG TACAACACGA
                                         3' UTR
         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  421  TTTATTGTAC TCCGCTTGTT ATCATTATAC CCACCTATGT TGCAACATCT TTTTGGATCC
       AAATAACATG AGGCGAACAA TAGTAATATG GGTGGATACA ACGTTGTAGA AAAACCTAGG
         PolyA tail
         ~~~~~~~~~~~~~~~
         3' UTR
         ~
  481  CAAAAAAAAA AAA      493
       GTTTTTTTTT TTT
```

FIG. 3

4 / 6

```
                            START
                            ~~~
                    M   S   Q   E   I   V   Q   S   G   Q   T   Y   I   I
  1  TCTCTTCCAG TTTCTACCAT GTCGCAAGAA ATTGTTCAAT CAGGACAAAC CTACATCATC
     AGAGAAGGTC AAAGATGGTA CAGCGTTCTT TAACAAGTTA GTCCTGTTTG GATGTAGTAG

     T   N   A   K   S   G   T   V   V   D   L   S   G   E   D   N   K   S   I   I
 61  ACTAACGCCA AATCCGGCAC AGTTGTTGAC CTTTCGGGCG AAGACAACAA ATCTATTATT
     TGATTGCGGT TTAGGCCGTG TCAACAACTG GAAAGCCCGC TTCTGTTGTT TAGATAATAA

     G   F   P   K   H   G   G   T   N   Q   R   W   T   L   N   W   T   G   K   S
121  GGATTCCCA AGCATGGAGG AACAAATCAG AGGTGGACCC TCAACTGGAC AGGGAAGAGT
     CCTAAAGGGT TCGTACCTCC TTGTTTAGTC TCCACCTGGG AGTTGACCTG TCCCTTCTCA

     W   T   F   R   S   V   S   S   E   M   Y   L   G   L   N   G   S   P   S   D
181  TGGACTTTCC GCTCCGTTTC TTCTGAAATG TATCTTGGCC TGAATGGCTC GCCGTCTGAT
     ACCTGAAAGG CGAGGCAAAG AAGACTTTAC ATAGAACCGG ACTTACCGAG CGGCAGACTA

     G   T   K   L   V   A   V   T   T   P   V   E   W   H   I   W   H   D   E   V
241  GGAACAAAAC TGGTAGCCGT GACCACCCCT GTTGAGTGGC ACATCTGGCA CGACGAAGTT
     CCTTGTTTTG ACCATCGGCA CTGGTGGGGA CAACTCACCG TGTAGACCGT GCTGCTTCAA

     D   P   S   T   Y   R   I   F   V   P   F   T   T   F   N   M   D   L   Y   A
301  GACCCTTCAA CTTATCGTAT CTTTGTACCT TTCACCACAT TCAACATGGA CCTCTACGCC
     CTGGGAAGTT GAATAGCATA GAAACATGGA AAGTGGTGTA AGTTGTACCT GGAGATGCGG
```

      A/G polymorphism   C/T polymorphism
          ~   ~
```
     Q   G   S   A   A   P   G   T   P   I   T   T   W   Y   T   W   K   G   I   H
361  CAAGGTAGTG CCGCCCCTGG TACGCCTATC ACAACTTGGT ATACATGGAA GGGCATCCAC
     GTTCCATCAC GGCGGGGACC ATGCGGATAG TGTTGAACCA TATGTACCTT CCCGTAGGTG
```

                    G/T polymorphism
                         ~
                    STOP
                    ~~~
```
     Q   T   W   R   F   E   L   A   *
421  CAAACGTGGA GGTTTGAACT AGCTTAGGGT CAGGTTTCGG ATGTAATTTG T   491
     GTTTGCACCT CCAAACTTGA TCGAATCCCA GTCCAAAGCC TACATTAAAC A
```
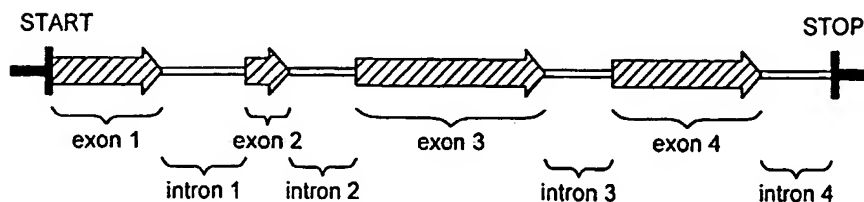
# FIG. 4



FIG. 5

# FIG. 6

```
START
~~ ~
                                    exon 1
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        M   S   Q   E   I   V   Q   S   G   Q   T   Y   I   I
  1   TCTCTTCCAG TTTCTACCAT GTCGCAAGAA ATTGTTCAAT CAGGACAAAC CTACATCATC
      AGAGAAGGTC AAAGATGGTA CAGCGTTCTT TAACAAGTTA GTCCTGTTTG GATGTAGTAG


                           exon 1
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                          intron 1
                                                          ~~~~~

        T   N   A   K   S   G   T   V   V   D   L   S   G   E   D   N   K   S
 61   ACTAACGCCA AATCCGGCAC AGTTGTTGAC CTTTCGGGCG AAGACAACAA ATCTAGTAAG
      TGATTGCGGT TTAGGCCGTG TCAACAACTG GAAAGCCCGC TTCTGTTGTT TAGATCATTC

                           intron 1
      ~~~~~~
121   TCGTTTTTAG TCCCATGTTT TTTTTTGTCA AAAAAAATTG ACTGACATAT TTTGTCTCCA
      AGCAAAAATC AGGGTACAAA AAAAAACAGT TTTTTTAAC TGACTGTATA AAACAGAGGT

                  exon 2
      ~~~~~~
      intron 1   intron 2
      ~   ~~~~~~~~~~~~~~~~~~~~~~~~~~~
        I   G   F   P   K   H   G   G   T   N   Q   R
181   GTTATTGGAT TTCCCAAGCA TGGAGGAACA AATCAGAGGG TAGGTCTAGA AATGCACCTC
      CAATAACCTA AAGGGTTCGT ACCTCCTTGT TTAGTCTCCC ATCCAGATCT TTACGTGGAG

                                               exon 3
                                     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                   intron 2
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                        W   T   L   N   W   T   G   K   S
241   GTTAATATTG GTTTTTATTG ACATTCATGA ACAGTGGACC CTCAACTGGA CAGGGAAGAG
      CAATTATAAC CAAAAATAAC TGTAAGTACT TGTCACCTGG GAGTTGACCT GTCCCTTCTC

                           exon 3
      ~~~~~~
        W   T   F   R   S   V   S   S   E   M   Y   L   G   L   N   G   S   P   S   D
301   TTGGACTTTC CGCTCCGTTT CTTCTGAAAT GTATCTTGGC CTGAATGGCT CGCCGTCTGA
      AACCTGAAAG GCGAGGCAAA GAAGACTTTA CATAGAACCG GACTTACCGA GCGGCAGACT

                           exon 3
      ~~~~~~
        G   T   K   L   V   A   V   T   T   P   V   E   W   H   I   W   H   D   E   V
361   TGGAACAAAA CTGGTAGCCG TGACCACCCC TGTTGAGTGG CACATCTGGC ACGACGAAGT
      ACCTTGTTTT GACCATCGGC ACTGGTGGGG ACAACTCACC GTGTAGACCG TGCTGCTTCA
```

6 / 6

```
        exon 3                          intron 3
    ~~~~~~~~~~~~~~~   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      D   P   S   T   Y
421 TGACCCTTCA ACTTATCGGT GAGTCCCCTA AATATTACTT GCTTGTGGTT CATACTAATA
    ACTGGGAAGT TGAATAGCCA CTCAGGGGAT TTATAATGAA CGAACACCAA GTATGATTAT


        intron 3                           exon 4
    ~~~~~~~~~~~~~~~   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                  I   F   V   P   F   T   T   F   N   M   D   L   Y   Q   G
481 CGTCGTTCGA AGTATCTTTG TACCTTTCAC CACATTCAAC ATGGACCTCT ACGCCCAGGG
    GCAGCAAGCT TCATAGAAAC ATGGAAAGTG GTGTAAGTTG TACCTGGAGA TGCGGGTCCC


                            exon 4
    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      S   A   P   G   T   P   I   T   T   W   Y   T   W   K   G   I   H   Q   T
541 TAGTGCCGCC CCTGGTACGC CTATCACAAC TTGGTATACA TGGAAGGGTA TCCACCAAAC
    ATCACGGCGG GGACCATGCG GATAGTGTTG AACCATATGT ACCTTCCCAT AGGTGGTTTG


                                      intron 4
                    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        exon 4
    ~~~~~~~~~~~~~~~~~
      W   R   F   E   L
601 GTGGAGGTTT GAACTAGGTA GGGCTTGCGA TCTCACCCGG ATCCTCCATG AACTAATGTG
    CACCTCCAAA CTTGATCCAT CCCGAACGCT AGAGTGGGCC TAGGAGGTAC TTGATTACAC


        intron 4           STOP
    ~~~~~~~~~~~~~~~~~~~  ~ ~~
661 ATCACGTCGT GTTCTAGCTT AGGTTCAGGT TTCGGATGTA ATTTGT            706
    TAGTGCAGCA CAAGATCGAA TCCAAGTCCA AAGCCTACAT TAAACA
```

FIG. 6 CONT'D